

Taller de R: Estadística y Programación

Taller 2

13/04/2021

En este taller se evalúan el módulo 3 del curso. Se presentan 2 tipos de taller (A y B), pero usted solo debe desarrollar 1 de ellos. Para realizar este trabajo, podrá hacer grupo de hasta tres personas. Sea creativo en su código (no hay una respuesta única, todos los métodos que permitan obtener la misma respuesta son validos). Cuando encuentre una ayuda en línea que le permita solucionar algún problema, no olvide citar la fuente. Por último, lea atentamente las instrucciones del taller.

Instrucciones

- No seguir las instrucciones tiene una penalización del **10%** de la nota total.
- Este taller representa el 30% de la nota del curso.
- El taller estará disponible en SICUA desde las 20:00 horas del 13 de abril de 2021.
- Al recibir el taller, uno de los integrantes del grupo debe crear un repositorio llamado **task-2** para almacenar las respuestas del taller. Este repositorio debe contener como mínimo las siguientes carpetas: **data**, **scr** y **views**. El usuario que crea el repositorio debe invitar como colaborador a los demás integrantes del grupo y al usuario [eduard-martinez](#).
- El taller debe ser terminado antes de las 23:59 horas del 28 de abril de 2021, no se calificará ningún commit que sea hecho después de esa fecha.
- Por favor sea lo más organizado posible y comente paso a paso cada línea de código, pero recuerden **NO** usar ningún acento o carácter especial dentro del código para evitar problemas al abrir los script en los diferentes sistemas operativos.
- En las primeras líneas del script debe escribir los nombres y los códigos de las personas que trabajaron en el código.
- Recuerde que debe elegir y desarrollar solo 1 taller (A o B).

Taller A

1. Loops

1.0. Para realizar este punto debe importar el archivo *data/output/lista.rds* en un objeto llamado *lista_df*. Esta lista contiene 74 dataframes con los de delitos cometidos en Colombia de 2010 a 2019.

1.1. Use un loop para eliminar las filas con NA y renombrar las columnas de cada elemento de la lista con los nombres que se encuentran en una de las filas del dataframe. Además cambie todos los nombres de las variables a minúsculas.

1.2. Asegúrese de crear una variable *tipo_delito* que almacene le tipo de delito.

1.3. Use la función `rbindlist` de la librería `data.table` para crear un dataframe que contenga todos los elementos de la lista. Asegúrese de llamar a este objeto `df`.

2. Familia `apply()`, `lapply()` y `sapply()`

2.1. Use la función `lapply` para pintar sobre la consola una tabla de frecuencia para cada variable del objeto `df`.

3. Funciones

3.1. Cree una función que convierta en minúsculas todos los elementos de un vector de caracteres (ojo solo si el vector es de caracteres). **Hint:** Use la función `is.character` y la función `tolower`.

3.2. Aplique la función anterior a todas las variables de `df` (puede usar un loop o la función `lapply`).

Taller B

1. Loop y función que limpia archivos

1.0. Para realizar este punto debe importar el archivo `data/output/lista.rds` en un objeto llamado `lista_df`. Esta lista contiene 74 dataframes con los de delitos cometidos en Colombia de 2010 a 2019.

1.1. Debe crear una función que lea cada elemento de la lista y elimine las filas con NA y renombre las columnas de cada elemento de la lista con los nombres que se encuentran en una de las filas del dataframe. Además cambie todos los nombres de las variables a minúsculas. La función debe crear para cada elemento una variable `tipo_delito` que almacene el tipo de delito.

1.2. Use un loop para aplicar la función sobre cada elemento del objeto `lista_df` y cree un nuevo objeto llamado `lista_clean`.

1.3. Use la función `rbindlist` de la librería `data.table` para crear un dataframe que contenga todos los elementos de `lista_clean`. Asegúrese de llamar a este objeto `df`.

2. Familia `apply()`, `lapply()` y `sapply()`

2.1. Use la función `lapply` para pintar sobre la consola una tabla de frecuencia para cada variable del objeto `df`.

3. Lapply

3.1. Use `lapply` para aplicar la función que creo en el punto **1.1.** sobre el objeto `lista_df`.

3.2. Compare los tiempos entre el loop (punto **1.2.**) y el `lapply`. Cuál es más rápido? **Hint:** use la función `Sys.time` para comparar los tiempos de ejecución.